SYSTEM AND METHOD FOR PERFORMING ELECTRONIC INFORMATION RETRIEVAL USING KEYWORDS

Abstract

Output documents similar to an input document are identified. A query is formulated using a list of best keywords from the input document to search for a first set of output documents. The list of best keywords is defined with a maximum number of keywords less than the total number of keywords in the list of best keywords that are identified as belonging to a domain specific dictionary of words and as having no measurable linguistic frequency. Lists of keywords are identified for each output document in the first set of documents. A second set of similar documents is determined using a measure of similarity that is computed between keywords identified in the input document and each output document in the first set of documents.